Top 50 Machine Learning Interview Questions & Answers

1) What is Machine learning?

Machine learning is a branch of computer science which deals with system programming in order to automatically learn and improve with experience. For example: Robots are programed so that they can perform the task based on data they gather from sensors. It automatically learns programs from data.

2) Mention the difference between Data Mining and Machine learning?

Machine learning relates with the study, design and development of the algorithms that give computers the capability to learn without being explicitly programmed. While, data mining can be defined as the process in which the unstructured data tries to extract knowledge or unknown interesting patterns. During this process machine, learning algorithms are used.

3) What is 'Overfitting' in Machine learning?

In machine learning, when a statistical model describes random error or noise instead of underlying relationship 'overfitting' occurs. When a model is excessively complex, overfitting is normally observed, because of having too many parameters with respect to the number of training data types. The model exhibits poor performance which has been overfit.

4) Why overfitting happens?

The possibility of overfitting exists as the criteria used for training the model is not the same as the criteria used to judge the efficacy of a model.

5) How can you avoid overfitting ?

By using a lot of data overfitting can be avoided, overfitting happens relatively as you have a small dataset, and you try to learn from it. But if you have a small database and you are forced to come with a model based on that. In such situation, you can use a technique known as **cross validation**. In this method the dataset splits into two section, testing and training datasets, the testing dataset will only test the model while, in training dataset, the datapoints will come up with the model.

In this technique, a model is usually given a dataset of a known data on which training (training data set) is run and a dataset of unknown data against which the model is tested. The idea of cross validation is to define a dataset to "test" the model in the training phase.

6) What is inductive machine learning?

The inductive machine learning involves the process of learning by examples, where a system, from a set of observed instances tries to induce a general rule.

7) What are the five popular algorithms of Machine Learning?

a) Decision Trees

b) Neural Networks (back propagation)

c) Probabilistic networks

d) Nearest Neighbor

e) Support vector machines

8) What are the different Algorithm techniques in Machine Learning?

The different types of techniques in Machine Learning are

a) Supervised Learning

b) Unsupervised Learning

c) Semi-supervised Learning

d) Reinforcement Learning

e) Transduction

f) Learning to Learn

9) What are the three stages to build the hypotheses or model in machine learning?

a) Model building

b) Model testing

c) Applying the model

10) What is the standard approach to supervised learning?

The standard approach to supervised learning is to split the set of example into the training set and the test.

11) What is 'Training set' and 'Test set'?

In various areas of information science like machine learning, a set of data is used to discover the potentially predictive relationship known as 'Training Set'. Training set is an examples given to the learner, while Test set is used to test the accuracy of the hypotheses generated by the learner, and it is the set of example held back from the learner. Training set are distinct from Test set.

12) List down various approaches for machine learning?

The different approaches in Machine Learning are

a) Concept Vs Classification Learning

b) Symbolic Vs Statistical Learning

c) Inductive Vs Analytical Learning

13) What is not Machine Learning?

a) Artificial Intelligence

b) Rule based inference

14) Explain what is the function of 'Unsupervised Learning'?

a) Find clusters of the data

b) Find low-dimensional representations of the data

c) Find interesting directions in data

d) Interesting coordinates and correlations

e) Find novel observations/ database cleaning

15) Explain what is the function of 'Supervised Learning'?

a) Classifications

b) Speech recognition

c) Regression

d) Predict time series

e) Annotate strings

16) What is algorithm independent machine learning?

Machine learning in where mathematical foundations is independent of any particular classifier or learning algorithm is referred as algorithm independent machine learning?

17) What is the difference between artificial learning and machine learning?

Designing and developing algorithms according to the behaviours based on empirical data are known as Machine Learning. While artificial intelligence in addition to machine learning, it also covers other aspects like knowledge representation, natural language processing, planning, robotics etc.

18) What is classifier in machine learning?

A classifier in a Machine Learning is a system that inputs a vector of discrete or continuous feature values and outputs a single discrete value, the class.

19) What are the advantages of Naive Bayes?

In Naïve Bayes classifier will converge quicker than discriminative models like logistic regression, so you need less training data. The main advantage is that it can't learn interactions between features.

20) In what areas Pattern Recognition is used?

Pattern Recognition can be used in

a) Computer Vision

b) Speech Recognition

c) Data Mining

d) Statistics

e) Informal Retrieval

f) Bio-Informatics

21) What is Genetic Programming?

Genetic programming is one of the two techniques used in machine learning. The model is based on the testing and selecting the best choice among a set of results.

22) What is Inductive Logic Programming in Machine Learning?

Inductive Logic Programming (ILP) is a subfield of machine learning which uses logical programming representing background knowledge and examples.

23) What is Model Selection in Machine Learning?

The process of selecting models among different mathematical models, which are used to describe the same data set is known as Model Selection. Model selection is applied to the fields of statistics, machine learning and data mining.

24) What are the two methods used for the calibration in Supervised Learning?

The two methods used for predicting good probabilities in Supervised Learning are

a) Platt Calibration

b) Isotonic Regression

These methods are designed for binary classification, and it is not trivial.

25) Which method is frequently used to prevent overfitting?

When there is sufficient data 'Isotonic Regression' is used to prevent an overfitting issue.

26) What is the difference between heuristic for rule learning and heuristics for decision trees?

The difference is that the heuristics for decision trees evaluate the average quality of a number of disjointed sets while rule learners only evaluate the quality of the set of instances that is covered with the candidate rule.

27) What is Perceptron in Machine Learning?

In Machine Learning, Perceptron is an algorithm for supervised classification of the input into one of several possible non-binary outputs.

28) Explain the two components of Bayesian logic program?

Bayesian logic program consists of two components. The first component is a logical one ; it consists of a set of Bayesian Clauses, which captures the qualitative structure of the domain. The second component is a quantitative one, it encodes the quantitative information about the domain.

29) What are Bayesian Networks (BN) ?

Bayesian Network is used to represent the graphical model for probability relationship among a set of variables .

30) Why instance based learning algorithm sometimes referred as Lazy learning algorithm?

Instance based learning algorithm is also referred as Lazy learning algorithm as they delay the induction or generalization process until classification is performed.

31) What are the two classification methods that SVM (Support Vector Machine) can handle?

a) Combining binary classifiers

b) Modifying binary to incorporate multiclass learning

32) What is ensemble learning?

To solve a particular computational program, multiple models such as classifiers or experts are strategically generated and combined. This process is known as ensemble learning.

33) Why ensemble learning is used?

Ensemble learning is used to improve the classification, prediction, function approximation etc of a model.

34) When to use ensemble learning?

Ensemble learning is used when you build component classifiers that are more accurate and independent from each other.

35) What are the two paradigms of ensemble methods?

The two paradigms of ensemble methods are

a) Sequential ensemble methods

b) Parallel ensemble methods

36) What is the general principle of an ensemble method and what is bagging and boosting in ensemble method?

The general principle of an ensemble method is to combine the predictions of several models built with a given learning algorithm in order to improve robustness over a single model. Bagging is a method in ensemble for improving unstable estimation or classification schemes. While boosting method are used sequentially to reduce the bias of the combined model. Boosting and Bagging both can reduce errors by reducing the variance term.

37) What is bias-variance decomposition of classification error in ensemble method?

The expected error of a learning algorithm can be decomposed into bias and variance. A bias term measures how closely the average classifier produced by the learning algorithm matches the target function. The variance term measures how much the learning algorithm's prediction fluctuates for different training sets.

38) What is an Incremental Learning algorithm in ensemble?

Incremental learning method is the ability of an algorithm to learn from new data that may be available after classifier has already been generated from already available dataset.

39) What is PCA, KPCA and ICA used for?

PCA (Principal Components Analysis), KPCA (Kernel based Principal Component Analysis) and ICA (Independent Component Analysis) are important feature extraction techniques used for dimensionality reduction.

40) What is dimension reduction in Machine Learning?

In Machine Learning and statistics, dimension reduction is the process of reducing the number of random variables under considerations and can be divided into feature selection and feature extraction

41) What are support vector machines?

Support vector machines are supervised learning algorithms used for classification and regression analysis.

42) What are the components of relational evaluation techniques?

The important components of relational evaluation techniques are

a) Data Acquisition

b) Ground Truth Acquisition

c) Cross Validation Technique

d) Query Type

e) Scoring Metric

f) Significance Test

43) What are the different methods for Sequential Supervised Learning?

The different methods to solve Sequential Supervised Learning problems are

a) Sliding-window methods

b) Recurrent sliding windows

c) Hidden Markow models

d) Maximum entropy Markow models

e) Conditional random fields

f) Graph transformer networks

44) What are the areas in robotics and information processing where sequential prediction problem arises?

The areas in robotics and information processing where sequential prediction problem arises are

a) Imitation Learning

b) Structured prediction

c) Model based reinforcement learning

45) What is batch statistical learning?

Statistical learning techniques allow learning a function or predictor from a set of observed data that can make predictions about unseen or future data. These techniques provide guarantees on the performance of the learned predictor on the future unseen data based on a statistical assumption on the data generating process.

46) What is PAC Learning?

PAC (Probably Approximately Correct) learning is a learning framework that has been introduced to analyze learning algorithms and their statistical efficiency.

47) What are the different categories you can categorized the sequence learning process?

a) Sequence prediction

b) Sequence generation

c) Sequence recognition

d) Sequential decision

48) What is sequence learning?

Sequence learning is a method of teaching and learning in a logical manner.

49) What are two techniques of Machine Learning ?

The two techniques of Machine Learning are

a) Genetic Programming

b) Inductive Learning

50) Give a popular application of machine learning that you see on day to day basis?

The recommendation engine implemented by major ecommerce websites uses Machine Learning